STUDY ON MODELING AND FORECASTING OF MILK PRODUCTION IN INDIA

Prema Borkar

Gokhale Institute of Politics and Economics,

BMCC Road, Deccan Gymkhana,

Pune – 411004. Maharashtra, India.

ABSTRACT

The paper describes an empirical study of modeling and forecasting time series data of milk production in India. Yearly milk production data for the period of 1979-1980 to 2014-2015 of India were analyzed by time-series methods. Autocorrelation and partial autocorrelation functions were calculated for the data. The Box Jenkins ARIMA methodology has been used for forecasting. The diagnostic checking has shown that ARIMA (1, 1, 1) is appropriate. The forecasts from 2015-2016 to 2019-2020 are calculated based on the selected model. The forecasting power of autoregressive integrated moving average model was used to forecast milk production for five leading years. These forecasts would be helpful for the policy makers to foresee ahead of time the future requirements of milk production, import and/or export and adopt appropriate measures in this regard.

Key words: ACF - autocorrelation function, ARIMA - autoregressive integrated moving average, Milk production and PACF - partial autocorrelation function.

India continues to be the largest producer of milk in the world. It has achieved the distinction of becoming the highest milk producing nation in the world, moving from a state of dependency to that of self sufficiency. It is hard to believe that milk production, which was 17 million tones in 1950, has increased to 137.7 million tons in 2014. This was possible due to the implementation of a number of successful initiatives by Government of India, both at the Central and State level, as it realized the importance of self sufficiency. In the current year, milk products have become an important item in the export kitty of India. This represents sustained growth in the availability of milk and milk products for our growing population.

This has become an important secondary source of income for millions of rural families and has assumed the most important role in providing employment and income generating opportunities particularly for marginal and women farmers. The per capita availability of the milk has reached a level of 290 grams per day during year 2011-12, but it is still lower than the world average of 284 grams per day. Most of the milk is produced by small, marginal farmers and landless laborers. About 14.46 million farmers have been brought under the ambit of 1,44,168 village level dairy corporative societies up to March 2011. Government of India is making efforts for strengthening the dairy sector through various development schemes like Intensive Dairy Development Programme, Strengthening Infrastructure for Quality & Clean Milk Production, Assistance to Cooperatives and Dairy Entrepreneurship Development Scheme.

In this study, these models were applied to forecast the production of milk in India. This would enable to predict expected milk production for the years from 2016 onward. Such an exercise would enable the policy makers to foresee ahead of time the future requirements for milk production, import and/or export of milk thereby enabling them to take appropriate measures in this regard. The forecasts would thus help save much of the precious resources of our country which otherwise would have been wasted.

METHODOLOGY

Respective time series data for this study were collected from various Government Publications of India. Box and Jenkins (1976) linear time series model was applied. Auto Regressive Integrated Moving Average (ARIMA) is the most general class of models for forecasting a time series. Different series appearing in the forecasting equations are called “Auto-Regressive” process. Appearance of lags of the forecast errors in the model is called “moving average” process. The ARIMA model is denoted by ARIMA (p,d,q), Where,

"p" stands for the order of the auto regressive process,

"d" is the order of the data stationary and

"q" is the order of the moving average process.

The general form of the ARIMA (p,d,q) can be written as described by Judge, *et al*. (1988).

$$\Delta^d y_t = \delta + \theta_1 \Delta^d y_{t-1} + \theta_2 \Delta^d y_{t-2} + \text{-----} + \theta_p y_{t-p} + e_{t-l} \alpha\, e_{t-1} - \alpha_2\, e_{t-2}\ \alpha_q e_{\,t-2} \qquad (1)$$

Where,

Δ^d denotes differencing of order d,i.e., $\Delta y_t = y_t - y_{t-1}$,

$\Delta_2 y_t = \Delta y_t - \Delta_{t-1}$ and so forth,

$Y_{\,t-1}$ ----- y_{t-p} are past observations(lags),

δ, θ_1 -------- θ_p are parameters (constant and coefficient) to be estimated similar to regression coefficients of the Auto Regressive process (AR) of order "p" denoted by AR (p) and is written as

$$Y = \delta + \theta_1\, y_{\,t-1} + \theta_2\, y_{\,t-2} + \text{----------} + \theta_p y_{\,t-p} + e_t \qquad (2)$$

Where,

e_t is forecast error, assumed to be independently distributed across time with mean θ and variance $\theta_2 e$, e_{t-1}, e_{t-2} ------ e_{t-q} are past forecast errors,

α_1, ---------- α_q are moving average (MA) coefficient that needs to be estimated.

While MA model of order q (i.e.) MA (q) can be written as

$$Y_t = e_t - \alpha_1\, \alpha_{t-1} - \alpha_2 e_{t-2} \text{ -------------- } \alpha_q e_{t-q} \qquad (3)$$

The major problem in ARIMA modeling technique is to choose the most appropriate values for the p, d, and q. This problem can be partially resolved by looking at the Auto correlation function (ACF) and partial Auto Correlation Functions (PACF) for the series (Pindyk & Rubinfeld, 1991). The degree of the homogeneity, (d) i.e. the number of time series to be differenced to yield a stationary series was determined on the basis where the ACF approached zero.

After determining "d" a stationary series $\Delta d\ y_t$ its auto correlation function and partial autocorrelation were examined to determined values of p and q, next step was to "estimate" the model. The model was estimated using computer package "SPSS".

Diagnostic checks were applied to the so obtained results. The first diagnostic check was to draw a time series plot of residuals. When the plot made a rectangular scatter around a zero horizontal level with no trend, the applied model was declared as proper. Identification of normality served as the second diagnostic check. For this purpose, normal scores were plotted against residuals and it was declared in case of a straight line. Secondly, a histogram of the residuals was plotted. Finding out the fitness of good served as the third check. Residuals were plotted against corresponding fitted values: Model was declared a good fit when the plot showed no pattern.

Using the results of ARIMA (p,q,d), forecasts from 2016 up to 2020 were made. These projections were based on the following assumptions.

- Absence of random shocks in the economy, internal or external.
- Milk price structure and polices will remain unchanged.
- Consumer preferences will remain the same.

RESULTS AND DISCUSSION

Building ARIMA model for Milk Production data in India

To fit an ARIMA model requires a sufficiently large data set. In this study, we used the data for Milk production for the period 1979-1980 to 2014-15. As we have earlier stated that development of ARIMA model for any variable involves four steps: identification, estimation, diagnostic checking and forecasting. Each of these four steps is now explained for milk production.

The time plot indicated that the given series is nonstationary. Non-stationarity in mean is corrected through appropriate differencing of the data. In this case difference of order 1 was sufficient to achieve stationarity in mean.

The newly constructed variable X_t can now be examined for stationarity. The graph of X_t was stationary in mean. The next step is to identify the values of p and q. For this, the autocorrelation and partial autocorrelation coefficients of various orders of Xt are computed (Table 1). The ACF and PACF (fig 1 and 2) shows that the order of p and q can at most be 1. We entertained three tentative ARIMA models and chose that model which has minimum AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). The models and corresponding AIC and BIC values are

ARIMA (p, d, q)	AIC	BIC
1 0 0	187.05	190.11
1 0 1	153.22	157.80
1 1 1	103.60	108.09

So the most suitable model is ARIMA (1,1,1) this model has the lowest AIC and BIC values.

Model parameters were estimated using SPSS package. Results of estimation are reported in table 2. The model verification is concerned with checking the residuals of the model to see if they contain any systematic pattern which still can be removed to improve on the chosen ARIMA. This is done through examining the autocorrelations and partial autocorrelations of the residuals of various orders. For this purpose, the various correlations up to 16 lags were computed and the same along with their significance which is tested by Box-Ljung test are provided in table 3. As the results indicate, none of these correlations is significantly different from zero at a reasonable level. This proves that the selected ARIMA model is an appropriate model. The ACF and PACF of the residuals (fig 3 and 4) also indicate 'good fit' of the model.

The last stage in the modeling process is forecasting. ARIMA models are developed basically to forecast the corresponding variable. There are two kinds of forecasts: sample period forecasts and post-sample period forecasts. The former are used to develop confidence in the model and the latter to generate genuine forecasts for use in planning and other purposes. The ARIMA model can be used to yield both these kinds of forecasts. The residuals calculated during the estimation process, are considered as the one step ahead forecast errors. The forecasts are obtained for the subsequent agriculture year from 2015-16 to 2019-2020.

CONCLUSION

In our study, the developed model for milk production was found to be ARIMA (1, 1, 1). The forecasts of milk production, lower control limits (LCL) and upper control limits (UCL) are presented in Table 4. The validity of the forecasted values can be checked when the data for the lead periods become available. The model can be used by researchers for forecasting of milk production in India. However, it should be updated from time to time with incorporation of current data.

This paper discloses the production of milk from 1979-1980 to 2014-2015 and also shows the future movement. To formulate future development plan for milk production, it is essential to know the previous condition and also see the future trend. In this study, forecasting is done by using some sophisticated statistical tools so that the government and policy makers can easily realize about the future development of milk production and could take initiatives to improve the production.

References:

Bartlett. M.S., 1964, On The Theoretical Specification of Sampling Properties of Autocorrelated Time Series. *J. Roy. Stat. Soc.,* B 8,pp 27–41.

Box, G.E.P., and G. M. Jenkins, 1970, Time series analysis: forecasting and Control, Holden Day, San Francisco, CA.

Box. G.E.P., and G.M. Jenkins, 1976, Time Series Analysis: Forecasting and Control. Rev. Ed. San Francisco. Holden-Day.

Brockwell. P.J., and Davis, R. A, 1996, Introduction to time series and Forecasting, Springer.

Brown. R.G., 1959, Statistical Forecasting For Inventory Control. New York, McGraw-Hill,

Iqbal. N., K. Bakhsh, A. Maqbool and A. S. Ahmad, 2005, Use of the ARIMA Model for forecasting Wheat area and production in Pakistan, *Journal of Agriculture & Social Sciences*, 2, pp 120-122.

Jenkins. G. M. and D.G. Watts, 1968, Spectral Analysis and its Application, Day, San Francisco, California, USA.

Ljunge. G.M., and G.E.P. Box, 1978, On a Measure of Lack of Fit in Time Series Models, Biometrika, 65, pp67–72.

Kendall. M. G., and A. Stuart, 1966, The advanced theory of statistics. Vol. 3. Design and Analysis and Time-Series. Charles Griffin & Co. Ltd., London, United Kingdom.

Holt. C.C., F. Modigliani, J.F. Muth and H.A. Simon, 1960, Planning, Production, Inventores, and Work Force. Prentice Hall, Englewood Cliffs, NJ, USA.

Muhammad. F., M. S. Javed, and M. Bashir, 1992, Forecasting Sugarcane Production in Pakistan using ARIMA Models, Pak. J. Agri. Sci, 9(1), pp 31-36.

Makridakis, S, A. Anderson, R. Filds, M. Hibon, R. lewandowski, J. Newton, E. Parzen and R. Winkler, 1982, The Accuracy of Extrapolation (time series) methods: Results of a Forecasting Competition, Journal of Forecasting Competition. J. Forecasting, 1, pp 111–153.

Meese. R., and J. Geweke, 1982, A Comparison of Autoregressive Univariate Forecasting Procedures for Macroeconomic Time Series. Unpublished Manuscript, University of California, Berkeley, CA, USA.

Prindycke. R.S., and D.L. Rubinfeld, 1981, Econometric Models and Economic Forecasts, 2nd Ed. New York, McGraw-Hill.

Quenouille. M.H., 1949, Approximate Tests of Correlation in Time- Series. J. Roy. Stat. Soc., B11, pp 68–84.

Saeed. N., A. Saeed, M. Zakria, T. M. Bajwa, 2000, Forecasting of Wheat Production in Pakistan using ARIMA models, International Journal of Agriculture and Biology, 4, pp 352-353.

Yule. G.U., 1926, Why Do We Sometimes Get Nonsence-corrleations Between Times Series. A study in Sampling and the Nature of Series. J. Roy. Stat. Soc., 89,pp 1–69.

Yule. G.U., 1927, On a method of Investigation Periodicities in Disturbed Series, With Special Reference To Wolfer's Sunspot Number. Phill. Trans., A 226, pp 267–298.

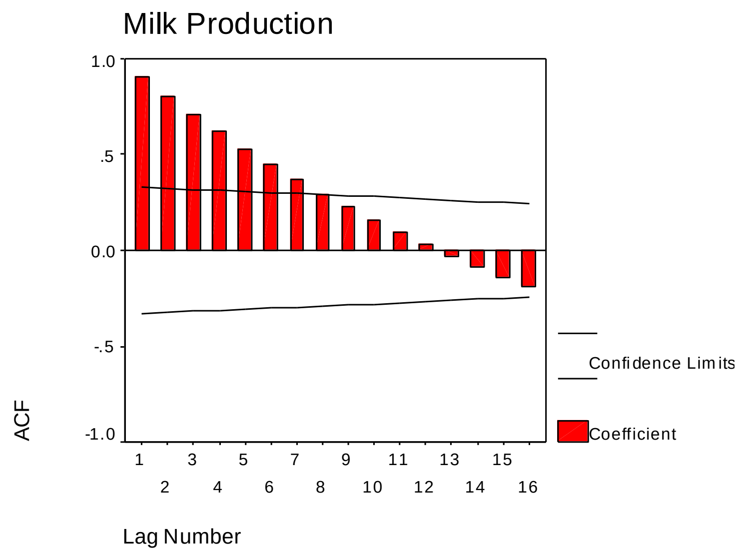

Figure 1: ACF of differenced data

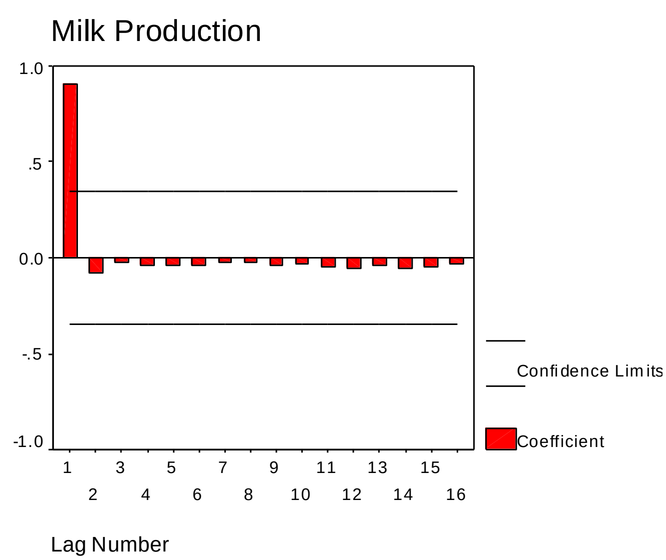

Figure 2: PACF of differenced data

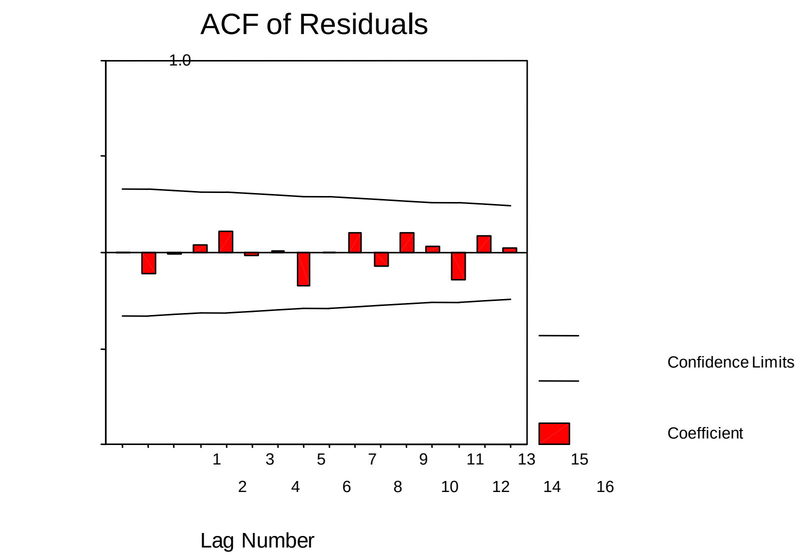

Figure 3: ACF of residuals of fitted ARIMA model

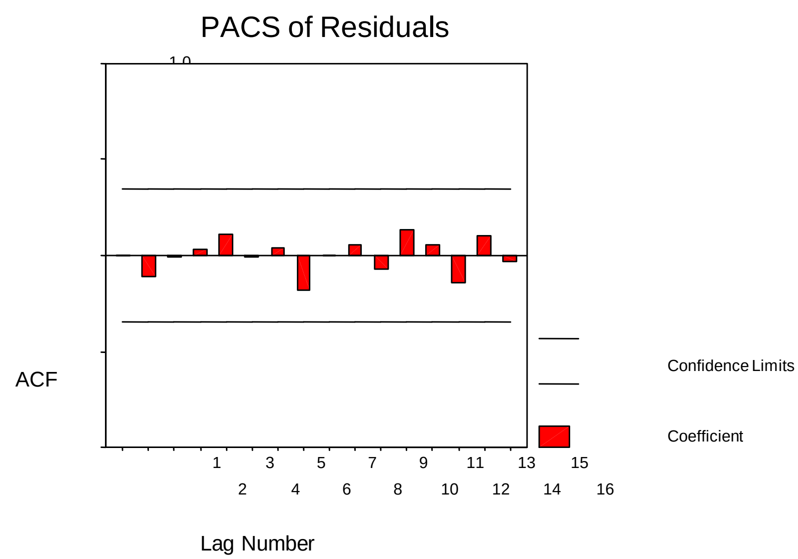

Figure 4: PACF of residuals of fitted ARIMA model

Table 1: Autocorrelations and partial autocorrelations

Lag	**Autocorrelation**	**Std.error**	**Lag**	**Partial Autocorrelation**	**Std.error**
1	0.905	0.164	1	-0.905	0.171
2	0.805	0.162	2	-0.079	0.171
3	0.711	0.159	3	-0.023	0.171
4	0.619	0.157	4	-0.040	0.171
5	0.531	0.154	5	-0.039	0.171
6	0.446	0.151	6	-0.039	0.171
7	0.367	0.149	7	-0.026	0.171
8	0.295	0.146	8	-0.023	0.171
	0.225	0.143	9	-0.041	0.171
10	0.159	0.140	10	-0.033	0.171
11	0.095	0.137	11	-0.051	0.171
12	0.031	0.134	12	-0.055	0.171
13	-0.029	0.131	13	-0.041	0.171
14	-0.087	0.128	14	-0.054	0.171
15	-0.141	0.125	15	-0.044	0.171
16	-0.190	0.121	16	-0.035	0.171

Table 2: Estimates of the fitted ARIMA model

		Estimates	Std Error	t	Approx sig
Non- Seasonal lag	AR1	0.8506264	0.19086189	4.4567643	0.00010745
	MA1	0.4685593	.34146201	1.3722150	.18017047
Constant		2.9643795	.60115745	4.9311200	.00002830
Number of Residuals		33			
Number of Parameters		2			
Residual df		30			
Adjusted Residual Sum of Squares		37.086014			
Residual Sum of Squares					
Residual Variance		1.2169926			
Model Std. Error		1.1031739			
Log-Likelihood		-48.802032			
Akaike's Information Criteria (AIC)		103.60406			
Schwarz's Bayesian Criterion (BIC)		108.09359			

Table 3: Autocorrelations and partial autocorrelations of residuals

Lag	Autocorrelation	Std.error	Box-Ljung	Df	Sig.	Lag	Partial Autocorrelation	Std. error
1	0.003	0.166	0.000	1.000	0.983	1	0.003	0.174
2	-0.0109	0.164	0.443	2.000	0.802	2	-0.109	0.174
3	-0.006	0.161	0.444	3.000	0.931	3	-0.006	0.174
4	0.039	0.158	0.506	4.000	0.973	4	0.028	0.174
5	0.111	0.156	1.013	5.000	0.961	5	0.111	0.174
6	-0.017	0.153	1.026	6.000	0.985		-0.011	0.174
7	0.011	0.150	1.032	7.000	0.994	7	0.036	0.174
8	-0.171	0.147	2.386	8.000	0.967	8	-0.180	0.174
9	-0.003	0.144	2.386	9.000	0.984	9	-0.004	0.174
10	0.102	0.141	2.907	10.000	0.984	10	0.055	0.174
11	-0.070	0.138	3.161	11.000	0.988	11	-0.069	0.174
12	0.103	0.135	3.744	12.000	0.988	12	0.136	0.174
13	0.032	0.132	3.802	13.000	0.993	13	0.056	0.174
14	-0.139	0.128	4.980	14.000	0.986	14	-0.140	0.174
15	0.084	0.125	5.428	15.000	0.988	15	0.101	0.174
16	0.025	0.121	5.472	16.000	0.993	16	-0.033	0.174

Table 4: Forecasts for Milk Production (2015-16 to 2019-2020)

(Million Tonnes)

Years	Forecasted yield	Lower limit	Upper limit
2015-2016	143.81	136.47	151.15
2016-2017	147.07	137.98	156.17
2017-2018	150.30	139.45	161.16
2018-2019	153.48	140.87	166.11
2019-2020	156.63	142.26	171.02